The Impact of Ohio's EdChoice on Traditional Public School Performance

Matthew Carr

Over the course of the last 35 years, traditional public school student achievement in the United States has been stagnant, despite myriad reform efforts and a doubling in total expenditures on K–12 education (Ravitch 2000, Hanushek 1986, Greene 2005). The ramifications of this academic achievement plateau on human capital development and thus the country's global economic standing are of paramount importance (Heckman and Masterov 2007). Thus, one of the most important public policy questions that government and society faces is how to improve the academic performance and quality of the nation's public education system.

Prominent among recent reform policies have been the twin pillars of accountability and school choice. Although many states were already implementing accountability programs, which required all schools to administer statewide standardized exams and then provided rewards and sanctions based on those results, the passage of the federal No Child Left Behind Act (NCLB) in 2001 made such policies a prerequisite for receiving federal funds. As a result, since the enactment of NCLB, all 50 states have now implemented statewide standardized testing and public reporting of school performance (Goertz 2005). The federal law provides for a series of sanctions for chronically underperforming schools, but states have

also supplemented the federal policy with rewards and sanctions of their own. Among the sanctions utilized by several states are requirements that students in failing schools be offered a wider range of school choices.

The State of Ohio, in particular, has pursued the education reform strategy of providing various school choice programs as a sanction for poor performance. Beginning in 1998, the state implemented a charter school program which is now tied directly to the accountability grading system. Similarly, in 2005 the state legislature enacted the Educational Choice Scholarship program (EdChoice), which provides vouchers to students in chronically underperforming schools, allowing them to attend private and religious schools.

School choice programs such as charter schools and vouchers serve two distinct purposes, at least in theory. The first purpose is to provide a direct and immediate outlet for students assigned to poorly performing traditional public schools. The second purpose, however, is perhaps the more important one. The competition that is created for students, as traditional public schools seek to maintain their enrollment levels (and the funding tied to those students) in the face of new schooling alternatives, is expected to create incentives for systemic improvements in the quality of education for all students (McEwan 2000, Chubb and Moe 1990). These hypothesized benefits of competition may or may not be realized when charter and voucher programs are actually implemented, and some argue that competition may be detrimental to the quality of traditional public schools (Fiske and Ladd 2000, Henig 1994, Meier and Smith 1995).

In this article, I evaluate the effects of the EdChoice voucher program on the academic performance of traditional public schools. Specifically, I investigate how exposure to the threat of losing students to the voucher program affected standardized test performance in traditional public schools. This measure does not take into account whether or not students participated in the EdChoice program, only whether schools faced the threat of potentially losing students. The study provides the opportunity to make two important contributions to the literature. First, it provides an analysis of a voucher program that has not yet been rigorously studied for its competitive effects on traditional public schools. Second, a novel approach is introduced to deal with the

potential confounding effects of stigma, or the motivation for schools to improve that may come with the receipt of a low grade publicly reported by the state.

To be clear, the study does not consider the effect of the EdChoice program on those students who actually use a voucher to attend another school. The participant effects literature suggests that such programs are typically beneficial for at least some subgroups of students.[1]

Theoretical Models of Choice-Induced Competition in K–12

The basic economic models of introducing expanded school choice through the use of vouchers focus on the expected market equilibrium that would result from alterations to demand and supply curves in schooling (McEwan 2000, Nechyba 2000, Epple and Romano 1998, Bradford and Shaviro 2000). In short, the implementation of a school voucher program is likely to lead to at least some shift in the schooling equilibrium along the dimensions of student and revenue reallocations. The effects of student sorting under a voucher program are important in three respects. First, vouchers may lead to increased segregation along any number of socioeconomic, demographic, or parental preference characteristics as students are self-selecting into various schools (Nechyba 2000). Second, student sorting may also alter the quality of peer groups within schools. Third, the sorting of students may also alter parental monitoring levels (McMillan 2000), as parents who are more likely to monitor the performance of their child's school are also more likely to exercise choice in a voucher program.

The schooling equilibrium is also greatly affected by the redistribution of resources among schools under a voucher program. Most state school funding mechanisms base funding levels, at least in part, on district enrollment figures. As a result, a traditional public school district may see revenues decline as students opt into voucher programs. This loss of revenue could have a detrimental impact on the traditional public schools, particularly if the funding

---

[1]See, for example, Cowen (2008); Greene (2001); Greene, Peterson, and Du (1999); Howell et al.(2002); Rouse (1998); Barnard et al. (2003); Peterson and Howell (2004); Wolf et al. (2010); Krueger and Zhu (2004); and Wolf (2008).

losses are greater than available cost reductions. However, it is also possible that the threat or actuality of losing resources could serve as a catalyst for operational reforms in the traditional public schools that increase efficiency, academic performance, or both. In this way, the reallocation of resources serves to potentially create competition between schools for students and the funding they represent.

The effects of the redistribution of students and resources created by a voucher program on the traditional public schools have been vigorously debated in the literature. Three general schools of thought have emerged: (1) Voucher programs lead to increased efficiency and academic performance for traditional public schools, as well as mitigating socioeconomic segregation levels (Friedman 1955, Chubb and Moe 1990, Hill and Guin 2002, Peterson 2006); (2) Voucher programs draw critical scarce resources (both dollars and motivated students and parents) away from the traditional public schools resulting in declining levels of quality, and exacerbate segregation along socioeconomic lines (Fiske and Ladd 2000, Belfield and Levin 2005, Henig 1994, Meier and Smith 1995); and (3) Voucher programs, in their current form, are too small and limited to have much real effect on the traditional public schools in terms of academic quality, efficiency, or segregation (Hess 2002, Hess and McGuinn 2002, Merrifield 2001). Because no one theory appears to be predominant among the research community, empirical evidence should be pursued where possible.

Competitive Effects Literature Review

In recent years, as the number of school choice programs has grown, so too has the amount of empirical research examining the competitive effects of school choice programs on student achievement (the effect of competition for students in traditional public schools). However, research on the competitive effects of publicly funded voucher programs remains limited. Most of these studies have examined the Milwaukee Parental Choice Program (MPCP), which allows disadvantaged students to use a voucher, or Florida's now defunct Opportunity Scholarship Program (OSP), which offers vouchers to students in failing schools. Two additional competitive effects studies have been conducted on the DC Opportunity

Scholarship Program and Florida's McKay Scholarship Program for Students with Disabilities.

The OSP has been by far the most studied voucher program in terms of its effects on traditional public school systems. The first set of studies focused on the initial implementation of the program in 1999 (Greene 2001; Greene and Winters 2003; Chakrabarti 2007a, 2008; Figlio and Rouse 2005). All of the studies examine changes in academic performance measures between a given pre-program period and a post period some number of years after implementation, usually with a difference-in-differences approach. Because the OSP creates a sample of schools that are either eligible for vouchers, or on the cusp of being eligible, or are not close to being eligible, researchers have typically sought out ways to compare these groups to distinguish the voucher threat effect.

Following the publication of the first study (Greene 2001), a number of important objections were raised, most prominently by Carnoy (2001). The two central issues he focused on were the potential bias introduced by regression to the mean and the confounding effects of stigma, or the "scarlet letter" (Carnoy 2001: 22) of receiving failing grades from the state. As a result, studies conducted after the Carnoy study have all attempted to test or correct for potential regression to the mean effects. Perhaps more importantly, later researchers have either conceded that they cannot disentangle voucher and stigma effects, and so simply reported a voucher/F-stigma effect (West and Peterson 2006, Chiang 2009), or have explicitly attempted to create models incorporating, and thus independently measuring, each effect (Rouse et al. 2007).

Changes made to Florida's school grading system between 2002 and 2003 set off a second round of evaluations which used those changes to create regression discontinuities (West and Peterson 2006, Rouse et al. 2007, Chiang 2009). All of these competitive effects studies conducted on the OSP in Florida have found positive effects, though the magnitude of those effects varies greatly. Also, some of the findings observed are attributed to a combined voucher/stigma effect, with a key exception being the work of Rouse et al. (2007), which found greater effects from the grade-only (stigma) period preceding the introduction of vouchers. Finally, despite the disadvantages of potential biases introduced by regression to the mean and the stigma confound, a significant advantage available to researchers of a failing school voucher program is the

ready availability of an exogenous policy shift that creates a sample of threatened and unthreatened schools.

The Milwaukee Parental Choice Program is a geographic-based, means-tested voucher program. As a geographic-based voucher program, which means that students must live within the district's boundaries to be eligible, researchers have typically sought out ways to compare more-threatened schools to less-threatened schools. To do this, they have compared schools with a larger number of eligible students, based on their income level, to those with fewer eligible students. Of the four studies evaluating competitive effects from the MPCP, three have found positive results (Hoxby 2002; Chakrabarti 2007b, 2008) and the fourth found no significant effect (Carnoy et al. 2007). The disadvantage of studying such programs is that every school in the district is subject to the program, and the number of eligible students in a particular school may be a poor proxy for differential pressure. In short, there is a greater likelihood for endogeneity to be an issue for the independent measure of competition.

Two final voucher competitive effects studies focus on the proximity of participating private schools and the number of participating private schools within a geographic radius as the measure of competition. One is a study of the DC Opportunity Scholarship Program, which is geographically based, and one is of the Florida McKay Scholarship Program for Students with Disabilities, which serves all students in the state with an Individualized Education Program (Greene and Winters 2006, 2008). The former study found no effects from the program on the traditional public school system, whereas the latter did find some positive effects based on disability type. The greatest obstacle to estimating accurately the impact of competition using measures such as the number of nearby schools is that they are endogenous. In short, the fact that private schools self-select their location leads to a non-random treatment of competition across traditional public schools. As a result, to use such measures would require highly sophisticated analytic techniques, such as instrumental variable regression, to produce unbiased findings.

Program Description

The original version of the Educational Choice Scholarship Pilot Program, enacted in June 2005, provided "14,000 students assigned to public schools that have been in academic emergency for three

consecutive years" to receive a voucher "worth up to $4,250 for kindergartners through eighth-graders and up to $5,000 for high-school students" (Smith-Richards 2005). However, just a year later, during the budget adjustment process in the spring and summer of 2006, the rules for program eligibility were amended again. This time, the number of eligible schools was increased by including those that had been in academic watch for three years straight as well (Smith-Richards 2006). The result was that students attending schools receiving either a rating of academic emergency (the equivalent of an F) or academic watch (the equivalent of a D) for the previous three consecutive years would be eligible to use an EdChoice voucher to escape their persistently failing schools. Thus, when the program finally went into effect in the fall of 2006, there were a number of schools that had been anticipating inclusion on the eligibility list, but others (those with some combination of D and F grades in the previous three years) that were quite surprised by the last-minute changes to the criteria.

In the fall of 2006, there were 2,193 students participating in the program from among the 99 traditional public schools that met the new eligibility criteria (Friedman Foundation 2007). In response to the relatively low take-up rate in the first year of the program, the state legislature again amended the eligibility criteria during the final full legislative session in December 2006. Under the revised rules, schools that received a rating of academic emergency (the equivalent of an F) or academic watch (the equivalent of a D) for any two of the previous three years would have students eligible to use a voucher (Stephens and Marshal 2006). The end result was that the number of schools where students could participate in the program increased dramatically. In the 2007–08 school year, more than 6,800 students were enrolled in private schools while using a voucher supplied by the EdChoice program (Friedman Foundation 2008).

Because of the multiple changes made to the eligibility criteria for schools, there are actually three distinct treatments that have occurred in the EdChoice program. The first treatment occurred in the 2005–06 school year, when schools that had received two consecutive academic emergency (F) ratings in the previous two years essentially faced the threat of being included in the program if they did not get at least an academic watch (D) rating that spring. At the time students were taking the state's standardized

263

exams in the spring of 2006, schools would not have known that an academic watch (D) rating would not keep them off the EdChoice eligibility list.

In the 2006–07 school year there was a somewhat larger number of schools facing the direct treatment of having students eligible to participate in EdChoice. This second treatment group no longer faces just a conditional threat, as it was the previous year when one more failing grade triggered the EdChoice sanction, but a direct threat where students could actually enter the program.

The third treatment group consists of those schools that had attained at least a continuous improvement (or C grade) rating during one of the previous three years, but otherwise had either D or F grades. These schools would have believed that their C grade (or better) precluded their facing the EdChoice sanction, only to be surprised by the criteria changes in early 2007 that suddenly made failure in just two out of the three previous years sufficient.

To provide some context for voucher-threatened schools, they come primarily from seven of the eight large urban school systems (Akron, Canton, Cincinnati, Columbus, Dayton, Toledo, and Youngstown). Cleveland was excluded from the program because the city had a pre-existing voucher program. The remaining eligible schools are typically found in the outer rings around the eight cities. Not all school systems experienced actual losses from the program. In the 2007–08 school year, five of the 31 districts with at least one eligible school building did not have a single student use a voucher. In fact, two-thirds of all students using a voucher came from just four school systems that year—Cincinnati, Columbus, Dayton, and Toledo. Looking at the question in terms of the marginal financial loss faced by districts with at least one EdChoice eligible school, the numbers are quite small. The average marginal loss was 0.81 percent of the total operating budget. Among the four systems with the largest number of EdChoice users, the average marginal funding loss amounted to 1.44 percent of the total operating budget. However, at an absolute level, the figures are much more striking. Among all eligible districts, total funding losses ranged from just $4,000 all the way up to almost $5.9 million.

Data and Methods

The data in this analysis were drawn from the Ohio Department of Education's publicly available data warehouse. Specifically, all data were downloaded from the department's "interactive Local Report Card" website. The data downloader provided access to highly disaggregated building-level information for every school in the state from the 2002–03 school year to the 2007–08 school year. The most significant limitation of using aggregate, building-level data is that alternative explanations for the results observed cannot be definitively ruled out (National Research Council 2002). Without student-level data, stronger assumptions are required about the plausibility of relationships. Similarly, including control variables and proxies and running different specifications of the underlying model are required to reduce the probability of alternative explanations for the relationship between voucher threat and performance changes.

Because the state has phased in a number of standardized exams since the passage of the federal No Child Left Behind Act, only four tests were available across the entire time period of interest: 4th and 6th grade reading and math exams. Three different outcome measures concerning student performance were collected: the percentage of students scoring at or above the proficient level; the percentage of students scoring in the limited proficiency level; and the percentage of students scoring in the advanced proficiency level. Limited proficiency is the lowest level of performance, and advanced proficiency is the highest. The reason for examining all three measures of student performance is that schools facing a voucher threat may behave strategically by focusing on so-called bubble students, those just above and below the proficiency cut-off, at the expense of higher and lower performers. However, schools facing a voucher threat may also decide to focus on those students most likely to use a voucher, the motivated high performers and the disaffected low performers. Because the highest and lowest performance categories do not represent actual scores, but rather the percentage of students who attain a particular score, ceiling and floor effects are less of a concern.

Also, a general performance indicator called the Performance Index Score (PIS) is used as a dependent variable. The PIS is calculated by taking the percentage of students at a given proficiency

level (from highest to lowest: advanced, accelerated, proficient, basic, below basic) and multiplying that percentage by a weight that is assigned to that level. The results are then summed at the building level with a maximum possible score of 120. Table 1 provides the descriptive statistics for the dependent variables.

The primary independent variable in these analyses is a dichotomous dummy (1/0) indicating whether a school faced the threat of vouchers in a particular year or not. As noted in the program description section, the nature of the threat is somewhat different in the 2006 school year than in the following two years. In the first year after the state legislature created the EdChoice program, schools faced only an indirect threat because they still had one year to improve scores enough to avoid eligibility. Put differently, if they had received an F the previous two years, their students would be eligible to use a voucher only if they also received an F again in the upcoming 2006 school year. As such, the threat of the EdChoice program in the 2006 year is defined as those schools that had already

TABLE 1
DESCRIPTIVE STATISTICS FOR DEPENDENT VARIABLES

Variable	Observations	Mean	Standard Deviation
Proficiency Rates			
4th Grade Reading	1,653 by 6 years	76.8	15.8
4th Grade Math	1,648 by 6 years	71.3	18.4
6th Grade Reading	1,086 by 6 years	75.0	16.7
6th Grade Math	1,086 by 6 years	68.4	19.7
Limited Rates			
4th Grade Reading	1,770 by 6 years	12.0	11.6
4th Grade Math	1,445 by 6 years	16.2	15.5
6th Grade Reading	1,640 by 6 years	19.4	17.9
6th Grade Math	1,303 by 6 years	23.8	20.4
Advanced Rates			
4th Grade Reading	1,685 by 6 years	9.5	7.3
4th Grade Math	1,845 by 6 years	20.2	13.7
6th Grade Reading	1,246 by 6 years	19.4	13.8
6th Grade Math	1,172 by 6 years	18.3	14.2
Performance Index	2,805 by 6 years	91.9	11.3

received an F grade in both of the previous two years. By contrast, for the 2007 and 2008 school years, schools faced the direct threat of students being able to leave with a voucher. Thus, in the final two years, voucher-threatened schools were those meeting the eligibility criteria and placed on the state's official list of schools where students could participate in EdChoice.

Furthermore, schools that faced a voucher threat were coded in a "once-always" manner. In short, once a school was placed on the voucher-threatened list, it was continually coded as a 1 in the following years whether the school continued to be on the official list or not. The reason for using this coding scheme was due to the assumption that a voucher-threatened school would make systemic changes in response and that to code a formerly threatened school as a 0 would fail to take into account the continuing effects of changes implemented in response to the original threat. Also, the once-always coding scheme is the most exogenous way to construct the variable. In each year after the program took effect, some schools could have been aware of their proximity to the voucher eligibility cut-off. By maintaining the coding after being placed on the eligibility list, confidence in the exogeneity of the variable is increased. The coding scheme leads to a sample of 72 schools that faced the threat of eligibility in the EdChoice program in the 2005–06 school year. In the 2006–07 school year there were 105 such schools, and the 2007–08 school year there were 112 threatened schools.

Finally, the reason for using the dichotomous construction of the voucher threat is that it avoids the endogeneity issues that accompany more continuous measures such as the number of eligible students or the proximity of private schools. Instead, this study examines only the threat that is imposed on a school either facing eligibility or actually on the eligibility list. This measure does not take into account whether or not students participated in the EdChoice program. The key feature is that the EdChoice program was designed and enacted in such a way that schools could not have foreseen the threat and did not play a role in shaping the eligibility rules. As such, the enactment of the law itself, and the amendments thereafter, constitute an exogenous policy shift or "shock" (West and Peterson 2006).

Because the EdChoice program is designed specifically as a sanction for chronic underperformance, it is crucial to include a

control variable for the quality of the schools in the sample. Otherwise, the estimates of the voucher threat treatment may be biased by the systematic application of the program to only the lowest-performing schools. If we believe there is something inherently different about an A school that distinguishes it from an F school, then it must be controlled for in the model. In this study I use a series of dummy variables to indicate the state rating grade (A = excellent, B = effective, C = continuous improvement, D = academic watch, and F = academic emergency) received by each school. The dummy variable for an F grade was chosen as the excluded comparison group.

The reason for not using other potential proxies for school quality is that those readily available, such as class size, teacher credentials, or spending per pupil, have not shown consistent relationships with school performance (Hanushek 2003) and including them would unnecessarily reduce degrees of freedom without adding much explanatory power. Better measures of school quality such as teacher turnover or teacher effectiveness are not available during this time period. Though imperfect, the school grade variables serve to capture at least some of the differences in school quality that may exist between schools in the sample.

Similarly, shifting peer groups resulting from changes in the composition of treatment and control schools may affect school performance. To control for such potential peer group effects, I include independent variables for the proportion of students in each school that are white, have a disadvantaged designation, and have a disabled designation. Along with controlling for changing peer groups, these variables also serve to, at least partially, take into account potential compositional effects from students departing their traditional public school for a charter or voucher-accepting private school.

Another potential source of bias might be the possibility of confounding effects generated by simultaneous NCLB sanctions. Again, the EdChoice program is specifically targeted to students in chronically failing schools, which are precisely those schools that are also targeted by the federal law and the sanctions it imposes. As such, control variables were included as a series of dummies for each of the seven NCLB treatment designations (as used by the Ohio Department of Education in their data coding): OK, At Risk, Improvement Year 1, Improvement Year 2, Improvement Year 3,

Improvement Year 4, and Improvement Year 5+. The dummy variable for an OK designation was chosen as the excluded comparison group. While research has indicated that the first three sanction levels are generally ineffective in creating much incentive for schools to improve (West and Peterson 2005), there certainly is reason to believe that the more severe sanctions associated with NCLB may induce schools to make significant changes in how they operate. Thus, it is important to control for any potential effects generated by the federal law and its consequences.

It has been noted in much of the previous literature (see Carnoy 2001) that evaluations of the competitive effects of failing schools voucher programs may mistake voucher threat effects with the effect of the stigma a school faces simply by virtue of getting a failing grade from the state that is highly publicized. In an effort to avoid this problem and to provide additional evidence on the independent effect of stigma, I include a separate independent variable for the scarlet letter. Specifically, the variable for stigma is a dichotomous dummy for whether or not a school has received a failing grade (a D or an F) in both of the previous two years. This construction has several advantages, with the most important being that it allows the stigma variable to get a year ahead of the voucher threat treatment variable in the time series. A key assumption is that the marginal difference between the stigma of failing twice in a row and failing three times in a row (and so being eligible for the voucher threat) is negligible.

The models are analyzed using a school fixed-effects regression. The form of the models is provided below:

$$PR_{st} = \beta + \emptyset V_{st} + \delta_j T_j + \gamma_j S_{jst} + \lambda N_{st} + \alpha_s + \varepsilon_{st}$$

where PR is the proficiency passage rate at school (s) and time (t); β is the constant; V denotes whether a school faces a voucher threat; T is a series of control variables for time; S is a set of school characteristics, including stigma; N is a series of control variables for NCLB sanctions; α denotes school fixed effects; and ε is a stochastic error term.

Fixed-effects models provide estimates of the relationship between changes in independent variables, within cases over time, on the outcome of interest. The fixed-effects method is ideal for this research question because it allows for an analysis that compares the differences in the performance trends of a school before and after it

269

faces the voucher threat and then compares those differences between treatment and control groups (Wooldridge 2003). In short, it factors out any time-invariant characteristics of each school, particularly any unobservable factors, because it compares each school to itself over time. For this study, time-invariant characteristics that are controlled for by the model include the community environment and school culture, because they are less likely to be amenable to change immediately following the introduction of a voucher threat than teacher or administrator behaviors. Unobservable characteristics such as student or teacher motivation are also taken into account by the fixed-effects approach.

To test the robustness of the model, several iterations were run, with and without the covariates for stigma and NCLB sanctions. Running several variations of the model in this way also provides information about potential multicollinearity inefficiency in the coefficient estimates of the primary independent variable (voucher threat). Because there is reason to believe that the voucher threat, stigma, and NCLB sanction variables may be functions of the same underlying phenomenon (chronic poor performance), it is important to estimate several iterations of the model. Similar results were observed across the multiple specifications, which increases confidence that estimated effects are accurate. Also, a bivariate correlation matrix of the independent variables is included in Table 2. None of the performance variables in the analytic model have a correlation greater than 0.41.

Additional analysis of smaller subsamples, such as comparing early threatened schools to those placed on the eligibility list in later years, were conducted to examine potential differential effects. Unfortunately, the relatively small number of voucher-threatened schools in such subgroups leaves the models insufficiently powered to detect any relationships between the dependent and primary independent variables.

One of the most salient threats to the validity of competitive effects estimates in a failing schools voucher program is regression to the mean. Chakrabarti (2007a) suggests that the best test for regression to the mean is to analyze the preprogram trends of to-be-treated and to-be-controlled groups. In other words, the performance trends of treated schools and nontreated schools are compared in the period before the program took effect. As such, regression analyses are run on the 2003 to 2005 performance

TABLE 2
BIVARIATE CORRELATION MATRIX

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1. Voucher Eligible	—											
2. GradeA	-0.08	—										
3. GradeB	-0.08	-0.50	—									
4. GradeC	0.02	-0.40	-0.41	—								
5. GradeD	0.14	-0.16	-0.16	-0.13	—							
6. Failure Stigma	0.40	-0.12	-0.12	0.02	0.20	—						
7. NCLB At Risk	-0.03	-0.17	-0.01	0.03	0.21	-0.07	—					
8. NCLB Yr1	-0.02	-0.13	-0.04	0.09	0.10	0.06	-0.11	—				
9. NCLB Yr2	0.07	-0.11	-0.07	0.08	0.15	0.26	-0.08	-0.05	—			
10. NCLB Yr3	0.19	-0.08	-0.06	0.04	0.14	0.19	-0.05	-0.04	-0.03	—		
11. NCLB Yr4	0.27	-0.07	-0.06	0.02	0.09	0.22	-0.04	-0.03	-0.02	-0.01	—	
12. NCLB Yr5	0.31	20.07	20.06	0.00	0.10	0.34	20.04	20.03	20.02	20.01	20.01	—

trends of (2006 to 2008) treatment and control schools for each of the academic performance variables used in the primary models. Specifically, the performance trend (the dependent variable) is operationalized as the difference between school performance in 2005 and 2003. OLS regressions were then run using the same set of covariates utilized in the primary analyses. Statistically significant results in the voucher threat variable would suggest the existence of regression to the mean and thus call into question whether results could truly be attributed to the competitive effects of the voucher program.

Results

Results of the analyses are presented and discussed below in four subsections: the analyses on proficiency passage rates and Performance Index Scores, on limited proficiency rates, on advanced proficiency rates, and regression to the mean tests. The primary results, the observed effects of the voucher threat and of stigma in the fully specified model, are presented in Table 3. Though not reported here, the demographic composition variables are all in the expected direction. Increases in the proportion of white students in the school population is associated with greater gains, while larger proportions of disadvantaged and disabled students are associated with slight, and usually statistically insignificant, declines.

Proficiency Passage Rates

Results in Tables 4 and 5 show that the voucher threat created by the EdChoice program is associated with a statistically significant increase in the proficiency rates of schools on the 4th grade reading tests. The voucher threat effect is associated with a gain of roughly 2.7 percentage points for the treatment schools in terms of proficiency passage rates. To put this in context, this effect represents an additional 2,150 students reaching the proficiency category. Of note, the stigma effect variable is not significant, which suggests that the scarlet letter of failure may not be spurring improvement in schools.

With the exception of the Performance Index Score (PIS), performance on the other three tests is not significantly associated with the voucher threat. For the two measures where significant results are observed (4th grade reading and the PIS), the outcomes

TABLE 3
VOUCHER THREAT AND STIGMA EFFECTS BY OUTCOME

Outcome Measure	Voucher Threat	Stigma
Proficient		
4th Grade Reading	0.17	0.00
4th Grade Math	0.00	0.00
6th Grade Reading	0.00	0.00
6th Grade Math	0.00	0.00
Limited		
4th Grade Reading	−0.16	0.30
4th Grade Math	0.49	−0.18
6th Grade Reading	0.30	0.00
6th Grade Math	0.23	−0.07
Advanced		
4th Grade Reading	0.38	0.36
4th Grade Math	0.30	0.00
6th Grade Reading	0.48	0.23
6th Grade Math	−0.43	−0.15
Performance Index Score	0.11	−0.04
Average	0.14	0.03

NOTES: Only statistically significant results (p<.10) are reported; insignificant results have a value of zero.

The direction of the *limited* effects have been switched for the purpose of calculating an average. As such, a reduction in the percentage of students scoring in the lowest category is a positive outcome and vice versa.

are positive, but have relatively small effect sizes. These results suggest that the voucher threat created by the EdChoice program may not have had much of an effect in terms of spurring schools to get more students over the proficiency benchmark. However, there is reason to believe that such students, so-called bubble students, may not be the most important focus for schools seeking to prevent losses to the voucher program. Instead, schools may have believed that the highest and lowest performers were more likely to utilize a voucher program. As such, the following sections turn to analyses of more disaggregated performance measures—the proportion of students scoring in the highest (advanced) and lowest (limited) proficiency categories.

TABLE 4
Grade 4 Reading Proficiency Level Analyses

Variable	Coefficient	Coefficient	Coefficient
Voucher Eligible	2.67***	2.83***	3.45***
Failure Stigma	−0.60	—	−0.29
At Risk (NCLB)	−1.56***	−1.42***	—
Improvement 1 (NCLB)	−2.57***	−2.35***	—
Improvement 2 (NCLB)	−0.77	−0.87*	—
Improvement 3 (NCLB)	0.03	−0.18	—
Improvement 4 (NCLB)	−1.06	−1.49	—
Improvement 5 (NCLB)	−0.46	−1.24	—
Constant	49.38	47.51	47.53
N	8,998	9,870	8,999
R^2-within	0.45	0.46	0.45

*Indicates $p<0.1$; ***indicates $p<0.01$.

Limited Proficiency Rates

The results from the 4th grade math limited proficiency analyses indicate strong gains made by voucher-threatened schools in reducing the percentage of students in the lowest performance category, as seen in Table 6. Interestingly, the stigma of failure also appears to be spurring improvement among schools, inde-

TABLE 5
Results of the Voucher Threat Variable on All Proficiency Level and PIS Analyses

Variable	Coefficient	P-value	Effect Size
4th Grade Reading	2.67	0.002	0.17
4th Grade Math	−0.58	0.570	—
6th Grade Reading	−0.59	0.576	—
6th Grade Math	−1.42	0.258	—
Performance Index Score	1.26	0.000	0.11

TABLE 6
GRADE 4 MATH LIMITED PROFICIENCY LEVEL ANALYSES

Variable	Coefficient	Coefficient	Coefficient
Voucher Eligible	−7.61***	−8.27***	−8.71***
Failure Stigma	−2.75***	—	−3.56***
At Risk (NCLB)	0.35	0.38	—
Improvement 1 (NCLB)	0.13	−0.16	—
Improvement 2 (NCLB)	−0.30	−1.36***	—
Improvement 3 (NCLB)	−4.17***	−5.30***	—
Improvement 4 (NCLB)	−3.25***	−4.38***	—
Improvement 5 (NCLB)	−2.85*	−5.10***	—
Constant	31.18	33.92	30.08
N	6,189	7,004	6,190
R^2-within	0.56	0.57	0.56

* Indicates $p<0.1$; *** indicates $p<0.01$.

pendently of the voucher threat effect, though to a much smaller degree. Of note, the NCLB sanctions beyond Improvement Year 1 (public school choice) also seem to be related to improvements in the percentage of low-performing students.

Examining the results of the voucher threat effect across the other three exams, as shown in Table 7, we see similarly strong gains in both 6th grade math and 6th grade reading. These effects are moderate in size, indicating gains between 0.2 and 0.5 standard deviations

TABLE 7
RESULTS OF THE VOUCHER THREAT VARIABLE ON ALL
LIMITED PROFICIENCY LEVEL ANALYSES

Variable	Coefficient	P-value	Effect Size
4th Grade Reading	1.89	0.001	0.16
4th Grade Math	−7.61	0.000	−0.49
6th Grade Reading	−5.28	0.000	−0.30
6th Grade Math	24.71	0.000	20.23

in the proportion of students scoring in the lowest level. Or, to put it another way, this represents a decline in the number of students scoring in the lowest category of between 3,800 and 6,000 students across the sample of threatened schools. However, in the analyses of the percentage of students with limited proficiency in 4th grade reading, the voucher threat is associated with an increase in the proportion of such students. The effect size is small, but still significant. Of note, the stigma variable is also positive and significant in the 4th grade reading analyses with a coefficient roughly twice as large as the voucher threat effect, suggesting that the scarlet letter has a damaging effect on school performance by increasing the number of students with limited proficiency in this case.

Advanced Proficiency Levels

The results of the analyses conducted on advanced proficiency levels indicate that high-performing students are not affected. Instead, the proportion of students scoring in this category actually increased. Strong gains are observed among voucher-threatened schools in increasing the proportion of students scoring in the highest proficiency level on the 4th grade reading exams. Similarly large gains are also associated with the failing grade stigma variable, independent of the voucher threat effects. Results are provided in Tables 8 and 9.

Across all four exams, analyses show gains in the proportion of students scoring in the advanced level, with one exception. The percentage of students scoring in the highest proficiency level on the 6th grade math tests appears to have declined moderately in response to the voucher threat. While the failing grade stigma variable also shows a negative relationship, the size of the effect is significantly smaller. Among those tests with a positive association between advanced proficiency levels and the voucher threat, the sizes of the effects are moderate, ranging from 0.3 to 0.4 standard deviations. These effects represent a gain of between roughly 2,200 and 5,300 students scoring in the highest category across the sample of threatened schools.

Regression to the Mean Tests

As previously noted, because the EdChoice voucher threat is, by definition, only faced by the lowest-performing schools, it is possi-

TABLE 8
GRADE 4 READING ADVANCED PROFICIENCY LEVEL
ANALYSES

Variable	Coefficient	Coefficient	Coefficient
Voucher Eligible	2.75***	3.31***	3.31***
Failure Stigma	2.63***	—	3.09***
At Risk (NCLB)	0.58**	0.42**	—
Improvement 1 (NCLB)	0.38	0.46	—
Improvement 2 (NCLB)	1.02**	1.26***	—
Improvement 3 (NCLB)	1.12	1.42**	—
Improvement 4 (NCLB)	0.54	0.49	—
Improvement 5 (NCLB)	4.03***	4.95***	—
Constant	6.17	4.67	6.84
N	8,034	8,856	8,035
R^2-within	0.39	0.40	0.39

* Indicates $p<0.1$; ** indicates $p<0.05$; *** indicates $p<0.01$.

ble that any gains observed are merely an artifact of a regression to the mean effect and not the result of competitive effects. To test for this potential bias, regression analyses were conducted on each of the exams (both proficiency and limited levels) to determine if pre-program trends existed between those schools that would ultimately be in the treatment and control samples after the program went into effect. The results of these tests are provided in Table 10.

TABLE 9
RESULTS OF THE VOUCHER THREAT VARIABLE ON ALL
ADVANCED PROFICIENCY LEVEL ANALYSES

Variable	Coefficient	P-value	Effect Size
4th Grade Reading	2.75	0.004	0.38
4th Grade Math	4.07	0.000	0.30
6th Grade Reading	6.58	0.000	0.48
6th Grade Math	−6.15	0.000	−0.43

TABLE 10
Regression to the Mean Tests

Variable	Coefficient	P-value
4th Grade Reading (P)*	0.13	0.929
4th Grade Math (P)	2.97	0.077
6th Grade Reading (P)	2.76	0.096
6th Grade Math (P)	0.76	0.676
Performance Index Score*	−3.08	0.231
4th Grade Reading (L)	4.66	0.000
4th Grade Math (L)*	0.35	0.787
6th Grade Reading (L)*	4.39	0.021
6th Grade Math (L)*	0.97	0.484

* Indicates a dependent variable where a significant effect was found from the voucher threat.

Several of the tests for regression to the mean did indicate significant pre-program trends between treatment and control schools. However, these results were found to exist only for dependent variables where a significant voucher threat effect was not observed. The 4th grade proficiency level score trends, in the pre-program, did not differ significantly from one another. Similarly, pre-program trends were not different between the groups on the Performance Index Score or 4th and 6th grade math limited proficiency levels. The trends observed in 6th grade limited reading levels are significantly different, but run in the opposite direction than would be expected if regression to the mean were present. In sum, the results observed in the analyses conducted on the competitive effects of the EdChoice program appear to be unbiased by potential regression to the mean effects.

Conclusion

The panoply of results observed in the preceding analyses present an interesting and compelling picture of the potential competitive effects of the EdChoice program. Perhaps the most notable findings are that the largest gains among the traditional public schools were observed in the highest and lowest categories of test performance. One hypothesis is that threatened schools chose to

focus most heavily on their highest and lowest performers, even though this led to little noticeable change to their overall proficiency passage rates. It has been argued that voucher-threatened schools would target so-called bubble students, those just above and below the proficiency cutoff scores, because overall passage rates are a key component of the federal Adequate Yearly Progress benchmarks and play a significant role in school grade determinations (Chakrabarti 2007a). Bubble students are also those for whom only small changes in test performance would have relatively large ramifications for a school's aggregate proficiency rates. Instead, we observe the exact opposite. The results suggest that voucher-threatened schools may be placing their focus on those students most likely to use the program to exit their residentially assigned school, those in the tails of the performance distribution.

It is also possible that schools focused on the highest and lowest performers because it is primarily the state's grading system, rather than federal AYP measures, that determine whether a school will face the voucher threat. Because the state places emphasis on both absolute proficiency passage rates and whether schools move students into the next highest proficiency category (i.e., getting students from limited to basic proficiency) there are incentives to target more than just bubble students. Had the state legislature decided to use the original eligibility criteria set forth by the governor in 2005, which was based solely on proficiency passage rates, the incentives would have been dramatically altered and the results may have conformed to the bubble student hypothesis. Of course, it is possible that schools respond to the immediate threat of losing students by focusing attention and resources on those perceived to be most likely to enroll in the program, and turn their attention to ways of getting off the eligibility list only as a secondary priority. These are explanatory hypothesis that future research, such as case studies or qualitative approaches, should explore by getting into the so-called black box of school operations to see if school leaders present internal strategies in direct response to a voucher threat.

In the end, the results of this study represent only the first step in a much longer process of discovery. Like other studies on this subject, an interesting and perhaps important relationship has been discovered between the implementation of a failing schools voucher program and subsequent changes in the performance of traditional public schools. However, a great deal of work remains to explore

possible causal explanations for the relationship observed. Finding answers to questions about the specific ways in which teachers, students, and administrators may have changed their behavior in response to a voucher threat or stigma, and about potential differential effects based on any number of contextual factors, is the next step for research on this subject.

References

Barnard, J.; Frangakis, C.; Hill, J.; and Rubin, D. (2003) "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association* 98 (462): 299–323.

Belfield, C., and Levin, H. (2005) "Vouchers and Public Policy: When Ideology Trumps Evidence." *American Journal of Education* 111: 548–67.

Bradford, D., and Shaviro, D. (2000) "The Economics of Vouchers." In C. Steuerle, V. Ooms, and G. Peterson (eds.) *Vouchers and the Provision of Public Services*. Washington: Brookings Institution Press.

Carnoy, M. (2001) "Do Vouchers Improve Failing Public Schools?" In M. Carnoy (ed.) *School Vouchers: Examining the Evidence*. Washington: Economic Policy Institute.

Carnoy, M.; Adamson, F.; Chudgar, A.; Luschei, T. F.; and Witte, J. F. (2007) *Vouchers and Public School Performance: A Case Study of the Milwaukee Parental Choice Program*. Washington: Economic Policy Institute.

Chakrabarti, R. (2007a) "Vouchers, Public School Response, and the Role of Incentives: Evidence from Florida." Federal Reserve Bank of New York, Staff Report No. 306.

——————(2007b) "Can Increasing Private School Participation and Monetary Loss in a Voucher Program Affect Public School Performance? Evidence from Milwaukee." Federal Reserve Bank of New York, Staff Report No. 300.

—————— (2008) "Impact of Voucher Design on Public School Performance: Evidence from Florida and Milwaukee Voucher Programs." Federal Reserve Bank of New York, Staff Report No. 315.

Chiang, H. (2009) "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics* 931: 1045–57.

Chubb, J., and Moe, T. (1990) *Politics, Markets and America's Schools*. Washington: Brookings Institution Press.

Cowen, J. (2008) "School Choice as a Latent Variable: Estimating the 'Complier Average Causal Effect' of Vouchers in Charlotte." *Policy Studies Journal* 36 (2): 301–15.

Epple, D., and Romano, R. (1998) "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects." *American Economic Review* 88 (1): 33–62.

Figlio, D., and Rouse, C. (2005) "Do Accountability and Voucher Threats Improve Low-Performing Schools?" National Bureau of Economic Research, Working Paper No. 11597.

Fiske, E., and Ladd, H. (2000) *When Schools Compete: A Cautionary Tale*. Washington: Brookings Institution Press.

Friedman, M. (1955) "The Role of Government in Education." In R. Solo (ed.) *Economics and the Public Interest*. New Brunswick, N.J.: Rutgers University Press.

Friedman Foundation (2007) *The ABCs of School Choice: 2006–2007*. Available at www.edchoice.org/newsroom/ItemAbc.do?filterId=1.

_____ (2008) *The ABCs of School Choice: 2007–2008*. Available at www.friedmanfoundation.org/friedman/downloadFile.do?id=277.

Goertz, M. (2005) "Implementing the No Child Left Behind Act: Challenges for the States." *Peabody Journal of Education* 80 (2): 73–89.

Greene, J. (2001) "An Evaluation of the Florida A-Plus Accountability and School Choice Program." Manhattan Institute for Policy Research, Civic Report.

_____ (2005) *Education Myths*. Lanham, Md.: Rowman and Littlefield.

Greene, J., and Winters, M. (2003) "When Schools Compete: The Effects of Vouchers on Florida Public School Achievement." Manhattan Institute for Policy Research, Education Working Paper No. 2.

_____ (2006) "An Evaluation of the Effect of D.C.'s Voucher Program on Public School Achievement and Racial Integration

after One Year." Manhattan Institute for Policy Research, Education Working Paper No. 10.

_____ (2008) "The Effect of Special-Education Vouchers on Public School Achievement: Evidence from Florida's McKay Scholarship Program." Manhattan Institute for Policy Research, Civic Report No. 52.

Greene, J.; Peterson, P.; and Du, J. (1999) "Effectiveness of School Choice: The Milwaukee Experiment." *Education and Urban Society* 31: 190–213.

Hanushek, E. (1986) "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24: 1141–77.

_____ (2003) "The Failure of Input-based Schooling Policies." *Economic Journal* 113: F64–F98.

Heckman, J., and Masterov, D. (2007) "The Productivity Argument for Investing in Young Children." National Bureau of Economic Research, Working Paper No. 13016.

Henig, J. (1994) *Rethinking School Choice: Limits of the Market Metaphor*. Princeton, N.J. : Princeton University Press.

Hess, R. (2002) *Revolution at the Margins: The Impact of Competition on Urban School Systems*. Washington: Brookings Institution Press.

Hess, R., and McGuinn, P. (2002) "Muffled by the Din: The Competitive Noneffects of the Cleveland Voucher Program." *Teachers College Record* 104 (4): 727–64.

Hill, P., and Guin, K. (2002) "Baseline for Assessment of Choice Programs." In P. Hill (ed.) *Choice with Equity: An Assessment by the Koret Task Force on K–12 Education*. Stanford, Calif.: Hoover Institution Press.

Howell, W.; Wolf, P.; Campbell, D.; and Peterson, P. (2002) "School Vouchers and Academic Performance: Results from Three Randomized Field Trials." *Journal of Policy Analysis and Management* 21: 191–217.

Hoxby, C. (2002) "School Choice and School Productivity (or Could School Choice Be a Tide that Lifts All Boats?)." National Bureau of Economic Research, Working Paper No. 8873.

Krueger, A., and Zhu, P. (2004) "Another Look at the New York City School Voucher Experiment." *American Behavioral Scientist* 47 (5): 658–98.

McEwan, P. (2000) "The Potential Impact of Large-Scale Voucher Programs." *Review of Educational Research* 70 (2): 103–49.

McMillan, R. (2000) "Competition, Parental Involvement, and Public School Performance." *Proceedings of the Annual Conference on Taxation*: 150–56.

Meier, K., and Smith, K. (1995) *The Case against School Choice: Politics, Markets, and Fools*. Armonk, N.Y.: M. E. Sharpe.

Merrifield, J. (2001) *The School Choice Wars*. Lanham, Md.: Scarecrow Press.

National Research Council (2002) *Scientific Research in Education*. Washington: National Academy Press.

Nechyba, T. (2000) "Mobility, Targeting, and Private-School Vouchers." *American Economic Review* 90 (1): 130–46.

Peterson, P. (2006) "The Use of Market Incentives in Education." In P. Peterson (ed.) *Choice and Competition in American Education.* Lanham, Md.: Rowman and Littlefield.

Peterson, P., and Howell, W. (2004) "Efficiency, Bias, and Classification Schemes: A Response to Alan B. Krueger and Pei Zhu." *American Behavioral Scientist* 47 (5): 699–717.

Ravitch, D. (2000) *Left Back: A Century of Battles over School Reform.* New York: Simon and Schuster.

Rouse, C. (1998) "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics* 113 (2): 553–602.

Rouse, C.; Hannaway, J.; Goldhaber, D.; and Figlio, D. (2007) "Feeling the Florida Heat: How Low-Performing Schools Respond to Voucher and Accountability Pressure." National Bureau of Economic Research, Working Paper No. 13681.

Smith-Richards, J. (2005) "Heat Turned Up on Charters: Budget Takes Aim at Subpar Schools, Expands Vouchers." *Columbus Dispatch* (23 June): A1.

———— (2006) "Voucher Bill May Spell New Woes for District." *Columbus Dispatch* (16 March): B1.

Stephens, S., and Marshall, A. (2006) "Ohio House Gives Taft School Reforms: Bill Toughens Curriculum to Get into College." *Cleveland Plain Dealer* (30 December): A1.

West, M., and Peterson, P. (2006) "The Efficacy of Choice Threats within School Accountability Systems: Results from

Legislatively Induced Experiments." *Economic Journal* 116 (510): C46–C62.

Wolf, P. (2008) "School Voucher Programs: What the Research Says about Parental School Choice." *Brigham Young University Law Review* 2: 415–46.

Wolf, P.; Gutmann, B.; Puma, P.; Kisida, B.; Rizzo, L.; Eissa, N.; Carr, M.; and Silverberg, M. (2010) *Evaluation of the DC Opportunity Scholarship Program: Final Report*. U.S. Department of Education, Institute of Education Sciences. Washington: U.S. Government Printing Office.

Wooldridge, J. (2003) *Introductory Econometrics*. 2nd ed. Mason, Ohio: South-Western.